KU LEUVEN
FACULTY OF ECONOMICS AND BUSINESS

# (Un)conventional combinations: At the origins of breakthrough inventions

**Antonio Della Malva and Massimo Riccaboni**

MSI_1507

# (UN)CONVENTIONAL COMBINATIONS:

# AT THE ORIGINS OF BREAKTHROUGH INVENTIONS[1]

Antonio Della Malva

Dept. of Managerial Economics, Strategy and Innovation (MSI), K.U. Leuven, Leuven, Belgium

Laboratory of Innovation Management and Economics (LIME), IMT School for Advanced Studies, Lucca, Italy

antonio.dellamalva@kuleuven.be

Massimo Riccaboni

Laboratory of Innovation Management and Economics (LIME), IMT School for Advanced Studies, Lucca, Italy

Dept. of Managerial Economics, Strategy and Innovation (MSI), K.U. Leuven, Leuven, Belgium

massimo.riccaboni@imtlucca.it

October 2014

## *Abstract*

Impactful inventions carry forward combinations of components which depart from common practices: they remove restrictions in the knowledge space by breaking conventional rules. In this paper we present a novel measure of the extent to which combinations in the inventive process conform to established practices. We borrow an established approach from the literature on product market diversification and adapt it to measure how combinations are typical or unconventional. We find that most of the inventive activities are grounded in conventional efforts, with rare instances of unconventional connections. Unconventionality is more likely to occur with experience, in teams and in large organizations. Moreover, patents which cite a widespread spectrum of previous results have a higher chance to identify unconventional connections. We also observe that inventions carrying forward unconventional combinations are cited more by future patent applications than conventional inventions.



JEL: O31; O32; C81; D01

# 1. Introduction

Technical change has been unanimously recognized to be the main engine of long-term economic growth and societal progress (Schumpeter, 1939). Particularly some inventions are unshakably mentioned amongst the most fundamental achievements of human kind and responsible for shifts in the technological paradigms (Dosi, 1982). These inventions are customarily addressed as breakthrough or radical inventions as they overcome existing bottlenecks in technical development and pave the way for new technical developments.

Despite such a common understanding of the widespread importance and fundamental impact of breakthrough innovations, the mechanisms responsible for the generation of high-impact inventions have been treated as black box and explored mostly speculatively. A common assumption made in the literature is that impact, and particularly extreme impact, is a function of the newness underlying the inventive process, which is modeled both as a process of search and recombination (Fleming, 2001). By looking at the inventive process as one of recombination and reconfiguration of existing ideas, newness is determined by those inventive acts which embed unfamiliar, unconventional or unconventional combinations (Simonton, 1999). As the search process is usually local, the extent to which combinations are conventional or typical is in turn a function of the distance in the technological space.

In this paper we draw on the literature on recombinant search and conceptualize the origins of novelty in the inventive process as a function of the proximity of the elements constituting the invention (Stuart and Podolny, 1996). Drawing from the literature in product market diversification, we adapt the measure of relatedness in the product space to account for the distance between each element of the knowledge space. This measure is population based and, similarly to the concept of technological regime, reflects the current set of beliefs and understanding of the relational structure of the knowledge space (Nelson and Winter, 1982). We claim that this measure captures the extent of conventionality in the recombinant process (Section 2). From the literature on the origins of breakthrough or radical inventions, we survey the antecedents of unconventionality and use the predictions from this literature to propose a set of testable statements (Section 3). For the purpose, we take advantage of the patent dataset at USPTO (Lai et al., 2014) and use data on patents, and the assignment to patent classes contained therein, over more than two decades – i.e. between 1975 and 2000 – to construct our measure of conventionality (Section 4.1).

We show that most combinations are indeed conventional as they occur between elements which have been similarly combined in the past. In the same fashion, most inventions are conventional as they embed combinations which are overly related. Only a handful of combinations bring together components which are substantially far apart, and these unconventional recombinant efforts come about in very few inventive acts. Unconventionality is also positively related to the impact of the inventions in the technological realm (Section 4.2).

Our results are in line with a view of unconventionality as the result of wide search, which spans technical domains to incorporate principles and solutions from other realms; it rests on the use of recent technical solutions and results from the attempt of combining large number of components. The organizational controls shed more light by indicating that large and experienced teams are mostly responsible for unconventional combinations in the inventive process, whereas lonely inventors are at disadvantage. Once we include controls for the size of the applicant and the institutional origin of it, large technological size becomes a meaningful predictor of unconventionality in invention and “garage” inventors are more conventional (Section 4.3).

This work belongs to a recent stream of research which inquiries the origins of breakthrough inventions and scientific discoveries by means of large scale databases (i.e. Ahuja and Lampert, 2001; Arts and Veugelers, 2013; Dahlin and Beherens, 2005; Fleming et al. 2007; Kelley et al., 2013; Schilling and Greene, 2011; Schoenmakers and Duysters, 2011; Uzzi et al. 2013). Most of the studies above, however, focus on measures of outcome such as breakthrough or radical inventions, defined as highly-cited patents or scientific papers. In this paper instead we present an account of the underlying dimension which is responsible for the extraordinary impact of some inventions, i.e. unconventional combinations. Our measure is based on patent class memberships as opposed to measured based on backward citations (Dahlin and Berhens, 2005; Uzzi et al., 2013) which are more sensitive to changes in the composition of the patent universe as compared to patent classes, which proportionally vary in small portions and after substantial time. A more detailed discussion of the results of our paper can be found in the final section (Section 5).

## 2. Measuring (Un)Conventionality: some theoretical considerations and a measure

### 2.1 Locus of Search in the Recombinatorial Process

Scholars have identified different forms characterizing the process through which new knowledge is created: combination of new components created by the inventor, new recombinations of existing components, and reconfiguration of existing combinations

(Schumpeter 1939, Nelson and Winter 1982, Weitzman 1998, Henderson and Clark 1990, Fleming and Sorenson, 2001). Therefore, knowledge is generated by integrating new components within an established framework or by modifying the existing framework to accommodate new reconfigurations (Schilling and Phelps, 2007). Knowledge generation hence starts with the search of diffused knowledge components (Cohen and Levinthal, 1990; Rosenkopf and Neckar, 2001). The set of combinable components comprises all bits of knowledge which are potentially available: it can entail existing components, previously untried components, or new components created by the inventor.[2] Inventors are expected to operate with an extraordinary large number of possible components and possibly an infinite number of combinations: the incremental process on which the creation of innovation is based exponentially increases the number of possible combinations with which individuals should deal. To ease the search process, subjects are used either to take into account familiar components which are locally available for new combinations, or to implement earlier utilized combinations. The choice of the components is therefore usually based on their availability, proximity, and saliency according to the inventor's aims (Fleming 2001).

Inventors usually search in the vicinity of their competences (Dosi, 1988; Stuart and Podolny, 1996). They rely on existing and certain solutions, whose past use has proved successful to the purpose (Cyert and March, 1992). The type of recombinant effort that results from local searches is characterized by high search depth (Katila and Ahuja, 2002), as it is geared towards increasing the understanding of a limited set of relationships among components. The exploration of local and familiar domains of knowledge is likely to deliver incremental solutions as the combinatorial possibilities can quickly exhaust (Fleming, 2001).

Inventors therefore reproduce or incrementally alter existing combinations, preserving the actual framework of relations among components.

The existing framework in which the components are related determines the cognitive availability of the latter. As relationships are scrutinized and challenged, the framework in which they are established is reinforced. Agents thus develop expectations on the nature of the relationships among the components forming the knowledge space and tend to constrain themselves to search within the existing boundaries of extant problems (Finke, 1995 as in Schilling and Greene, 2011). The pattern of association of the components therefore reflects conventions and common understanding of the possible interdependencies.

The continuous exploitation of local reservoirs of knowledge can lead to inventive traps, where inventors find themselves trapped in inefficient local optima. Extending the breadth of the knowledge base from which components are sourced is expected to bring to outcomes with higher degree of novelty and originality (Levinthal and March, 1993; Fleming, 2001). The number of possible combinations used in an invention increases with the set of elements which are available to the inventor in the generative phase. Furthermore, the broader the search scope in the generation phase, the more likely that inventors will combine components which stand far apart in the technological space.[3] From a cognitive standpoint, being exposed to a variety of sources can lead agents to analyse the same problem for different angles, and hence re-conceptualize it so that new elements are integrated into an existing interpretative framework (see Schilling and Greene, 2011, for an overview). Tthe inclusion of novel elements in established interpretative frameworks challenges the existing cognitive structures and can lead to the generation of novel and overlooked combinations (Fleming, 2001; Simonton, 1999).

Combinations which relate components which are scarcely if at all used together are therefore unconventional or unconventional.

### 2.2 Measuring Conventionality

Following the discussion above, we build a measure of conventionality in recombinations. This measure should reflect the distance between elements in the space of components as a function of the commonalities shared by the components. From the literature on the diversification of the business activities within firms, we borrow the measure of relatedness and its conceptualization, used to describe the diversification of firms in the product market as proposed by Teece et al. (1994) and adapted to describe the diversification patterns of firms at the technological portfolio level (Breschi et al., 2003; Nesta and Saviotti, 2005).

Two elements constituting a diversified set, for instance two products or two technologies in the portfolio of a firm, are said to be related if their joint occurrence is not driven by a random process. This is usually the outcome of existing commonalities or synergies between the two elements. The concept of coherence extends the rationale behind relatedness to the whole set of elements to capture the systematic relatedness of the elements comprising it.[4]

We follow the same line of reasoning and measure the extent to which each pair of components constituting single recombinant acts are related to each other, that is close in the knowledge space. In line with the empirical literature on the determinants of inventive impact (Fleming, 2001; Dahlin and Beherens, 2005; Schoemakers and Duyster, 2010), we use patent documents and the occurrence of patent classes therein as base for the construction of the measure. A patent has membership in one or more patent subclasses which are the objects to be combined. The extent to which each possible pairwise combination of patent subclasses actually occurs within each patent determines the starting point for the calculation of the measure. Let $C_{iz} = 1$ if patent $z$ has membership in class $i$, and 0 otherwise. The number of patents having simultaneously membership in classes $i$ and $j$ is

$$J_{ij} = \sum_z C_{iz} C_{jz} .$$

Raw counts of the number of patents having membership in each pairwise subclass combination, however, cannot be taken directly as a measure of relatedness. Although $J_{ij}$ increases with the relatedness of $i$ and $j$, it also increases with $n_i$ and $n_j$, the number of patents having membership in each class of the couple. Thus, large values of $J_{ij}$ might simply reflect intense inventive activities in $i$ and $j$. Therefore, $J_{ij}$ must be adjusted for the number of patents that would have simultaneous membership both in $i$ and $j$ under the null hypothesis that classes were randomly assigned to inventions. Teece et al. (1994) show that the joint occurrence of two objects $i$ and $j$ follows an hypergeometric distribution against which the null hypothesis can be tested. Hence relatedness, $\tau_{ij}$, is measured as the difference between the observed pattern of co-occurrences of $i$ and $j$ and the expected one:

$$\tau_{ij} = \frac{J_{ij} - \mu_{ij}}{\sigma_{ij}}$$

Where $\mu_{ij}$ is equal to the expected number of patents with simultaneous membership in $i$ and $j$ under the observed occurrences of $i$ and $j$ and $\sigma_{ij}$ the standard deviation of the observed occurrence.[5] This measure thus reports the extent to which a combination of patent subclasses appears as novel or conventional. When this measure is large, inventors systematically combine $i$ and $j$ in their inventions and thus the two components are related in the technical space; on the opposite, when it takes values close to 0 or even negative, the measure indicates that unexpectedly few inventions embed the two components given their separate use; consequently *i* and *j* are unrelated and their joint use will be rather novel or unconventional.[6] The measure is also population-based, in the sense that it reflects the actual state of relationships between elements of the space at a given point in time, and built around the "principle of survival" as the actual configuration of interdependences among components is the result of successful attempts and consequently weak or nonexistent links represent overlooked connections or failed trials. This feature enables to delineate the actual boundaries of the conceptual space and consequently any act of modifying sensibly the latter at any time.

## 3. Sources of (Un)Conventionality

A growing empirical literature has analyzed high impact or breakthrough or radical inventions, detailing a variety of determinants of impact (i.e. Fleming, 2001; Kelley et al., 2013; Schilling and Greene, 2011; Schoenmakers and Duysters, 2010 among others). Although these studies only speculate on the role of novelty in the determination of highly impactful inventions, they advance arguments which mostly pertain the sources of novelty (or unconventionality as we prefer to define it).

One of the most discussed issues surveyed in the studies above is the extent to which unconventionality is the outcome of the recombination of existing knowledge or it relies on completely new solutions. A stream of literature has argued that novelty in the knowledge base used for the generation of inventions relies on completely new technical knowledge, hence not yet embedded in existing inventions (van de Poel, 2003). A second stream of research instead points to the role of existing components, and their recombinations, expectedly distributed across the technological space (See section 2). Under the first view, novelty is carried forward by little if not existent references to previous inventive efforts (Ahuja and Lampert, 2001); unconventional combinations instead might find their rationale in the scientific realm and find their way in the technological domain (Dahlin and Behrens, 2005). The second perspective instead posits that the knowledge base from which unconventional recombinations are sourced is broadly distributed. Despite being a repository of knowledge with potential technological implications (not yet exploited), Science can work as a map in the technological space, allowing inventors to move within the latter with greater foresight (Fleming and Sorenson, 2004). By elaborating and testing theories of general validity, Science helps predict the outcome of scarcely, if at all, tested combinations, guiding thus inventors in their search beyond the existing cognitive boundaries.

Despite the different realms comprising the knowledge space, proximity has been defined in variety of terms. The temporal dimension has recently gained noteworthy attention (Neckar, 2003). The debate revolves around the contribution of novel and emerging bodies of knowledge to the generation of original solutions as opposed to the contribution of more mature ones. Emerging technologies usually bring about novel solutions, embed an higher degree of novelty in the proposed solutions and hence expand the current space for recombinations – for instance by bringing to the market new components themselves (Ahuja and Lampert, 2001). Mature technologies, on the opposite, tend to be "… well understood and offer greater reliability relative to more recently developed and less tested" technologies (Ahuja and Lampert, 2001, p. 527). Hence, familiarity with the nature and properties of older technologies will be substantially higher. Unconventional recombinations are also expected to be the result of combinations of older and emerging knowledge bases. As they result from the association of distant bodies of knowledge, such recombinant efforts will most often link bodies of knowledge with high internal coherence – i.e. areas of the knowledge space whose existing interdependences are mostly understood – but loosely recombined among themselves. A useful analogy in this respect is the realm of Science, where new contributions bear a tension between conformity to the "currently predominant beliefs about the nature of things" (Polanyi, 1962: 58) and dissent from it.

The organizational literature has extended the discussion on the sources of impactful inventions to include the role of inventors and teams. The debate revolves around the role of teams in the process of idea generation and retention. The question at the core of the debate is whether teams facilitate the recombination of dispersed competences, distributed across team members (Singh and Fleming, 2010) or whether they generate frictions in the phase of retention of creative ideas (Paulus and Nijstad, 2003). Advocates of the latter view embrace the "myth of the lone inventor" as source of unconventional solutions because teams are plagued by collaborative frictions in the process of idea generation (Mullen et al. 1991). Proponents of the former view claim that collaboration enables greater combinational opportunities and that teams are better endowed in the "sorting and identification of most promising ideas" (i.e. Singh and Fleming, 2010, p.42). In this respect, inventors' experience plays a crucial role in that it determines the extent of combinatorial possibilities and the ability to select promising inventive venues (Fleming et al., 2007; Hargadon and Sutton, 1997; Schilling and Greene, 2011).

The debate on the origins of novel or unconventional inventions is also one of the cornerstones of the industrial organization discussion. Scholars have been debating as to whether the type of organization in which inventions occur - large firms vs small firms – has an influence on the extent of (un)conventionality in recombination in the inventive process. On the one hand, large firms are considered to be at disadvantage with the generation of unconventional solutions as they are trapped in established routines and product lines, around which new solutions are incrementally developed (Hill and Rothaermel, 2003). On the other hand, firms can be thought as repositories of knowledge and competences (Grant, 1996) whose potential for recombination depends directly on firm size. This assumption is consistent with theories of industry evolutions via corporate spin-offs, where unconventional ideas are rejected by incumbent firms because of mismatch with the firms' main strategy (Klepper and Thomson, 2010). Hence large firms are a seedbed for (un)conventional combinations, whose exploitation will depend on strategic decisions.

The discussion above leads us to the following testable statements:

a) Unconventionality is the result of searches which span distant domains and hence entails the use/recombination of a broad array of components/solutions;
b) Unconventionality is affected by the organizational structure, team/organization, in which search occurs;
c) Unconventional combinations contribute to overcoming inventive traps and hence are related to higher inventive impact.

# 4. Results

## 4.1. Data

We use U.S. patent data from 1975 to 2000 inclusive (Lai et al., 2014), to measure the originality of inventive outputs. In line with most of research on patent data (Hall et al., 2001), only the utility patents were used. [7] The unit of analysis in the resulting models is the individual patent. The information contained in patents enables to model the extent to which the components used in the generation of the invention are combined in an unconventional fashion. In particular, we used detailed information about the patent's technology in class and subclass references (there are over 400 classes, and over 100,000 subclasses). Classes reflect broad technological areas, whereas subclasses reflect specific technological components within a given technological area. Most critical to this study is the listing of the technological components used in the generation of the invention and their joint occurrence across the whole universe of patents at the USPTO.

Aside from containing a great deal of technical information, a single patent also provides a rich amount of individual and organizational-level data about the individuals who worked on the invention. It contains the patent number, the date of application and grant, all inventors' names (also referred to as the authors) and hometowns, the assignee (i.e., the owner of the patent and typically identifies the organization for which the inventor works, such as a firm, a university or government, or the inventor himself).

### 4.2. Unconventional Combinations

Figure 1 reports the distribution of $\tau_{ij}$ for all patent subclass pairs as observed between 1980 and 2000. As the number of co-occurrences among patent subclasses can be highly volatile over time, we use 5 year moving averages. For the sake of exposition, we display the natural logarithm of $\tau_{ij}$. [8]

[FIGURE 1: THE DISTRIBUTION OF $T_{ij}$ ACROSS PAIRS]

The figure clearly shows that most combinations are highly conventional; only a handful of them show values of $\tau_{ij}$ which are close to the zero, and are hence original or unconventional. For instance, among the most unconventional combinations we can find the attempts to explore biotechnology-related applications in the late 1990s. The patent subclass *435/320.1* [Molecular Biology (435); Vector, per se (e.g., plasmid, hybrid plasmid, cosmid, viral vector, bacteriophage vector, etc.) (320.1)] [9] appears to be combined in an unconventional fashion with *425/401* [Drug (425); Cosmetics, antiperspirants, dentifrices (401)], and 707/3 [Data Processing: Database and File Management, Data Structures, or Document Processing (707); Query processing (i.e. searching) (3)]. [10] The two examples document the attempts to explore new applications for the nascent biotechnology sector: the first is the application of genetic engineering to the domain of cosmetics, whereas the second relates to bio-informatics.

Similarly, we can derive patent-based measures of conventionality, on the basis of the distributional properties of $\tau$ for each pairwise combination of patent subclasses within each patent. To this purpose, we provide two indicators of the degree of conventionality in an invention: the median and the minimum value of $\tau$ among the possible pairwise combinations contained in an invention. The median captures the degree of conventionality around the main bulk of combinations within the invention, whereas the minimum value indicates the least conventional recombinant act within an invention. Most patents embed a high degree of conventionality in the combination of their constituent parts. More than half of the patents in the sample have a median $\tau$ larger than 33, whereas only 28 patents have a median $\tau$ below 0. When we look at the minimum value of $\tau$ within each patent, more than half of the patents combine components whose $\tau$ is above 17; the occurrence of negative values is a rare event as well. All in all, the preliminary evidence provided so far indicates that the inventive process relies mostly on conventional recombinations and only rarely embed efforts which are unconventional.

Table 1 and table 2 report the distribution of the median tau respectively across years of application and technology domain of the focal invention. On average, inventions are less conventional over time; yet, over time there is a tendency to both exploit established trajectories and to move beyond the existing boundaries as we also observe that the dispersion of conventionality increases over time. Table 2 provides further evidence on the goodness of our measure, suggesting that inventions in domains like “Apparel and Textile” and “Furniture, House Fixtures” are more conventional than ICT related inventions like “Semiconductors” or “Computers”, which for instance find applications in a multitude of other domains.

[TABLE 1: DISTRIBUTION OF CONVENTIONALITY ACROSS YEARS]

[TABLE 2: DISTRIBUTION OF CONVENTIONALITY ACROSS TECHNOLOGIES]

We present the distribution of forward citations, corrected for scope of the patent, year and technology, by typology of invention.[11] We differentiate according to the extent of conventionality, setting a threshold at the $10^{th}$ centile of both the median and the minimum level.

Figure 2 reports the expected number of forward citations received by the focal inventions according to the classification as above. Inventions combining components in an unconventional fashion are on average more cited.[12] The effect is more pronounced for those inventions which are unconventional in their most unconventional recombination as compared to inventions which are unconventional at the core of their combinations. Furthermore, there is evidence that the inventions which combine unconventional combinations within an established framework enjoy the highest impact. This result is in line with the results by Schilling and Greene (2011), who argue that it suffices a very small amount of unconventional combinations to connect large and established bodies of knowledge, otherwise distant.

[FIGURE 2: FWD CITATIONS BY TYPOLOGY OF INVENTION]

Finally, we investigate the evolution of conventionality across combinations. We are interested in understanding the within-combination variation in the degree of conventionality over time. Table 3 displays the results of a fixed effect panel regression of our measure of conventionality against a trend variable and a set of year dummies. The aggregated results suggest a negative correlation between the trend variable and conventionality. As we already observed in Table 1 for inventions, there is a tendency to combine components in an unconventional manner. We then allow the change in conventionality to vary across different levels of initial conventionality, that is conventionality observed in the first instance of a combination in the database. With the sole exception of the portion of the distribution between the 25$^{th}$ centile and the median, all coefficients for the time trend are statistically significant. The results indicate that a move toward lower levels of conventionality is occurring in the central part of the distribution. On the opposite, conventionality increases for extreme values of initial conventionality: highly unconventional combinations become more conventional, at a faster rate than more conventional ones become unconventional, and conventionality strengthens over time for highly conventional combinations. Unconventional combinations are therefore at the core of the reconfiguration of the current cognitive framework. In a tension between conformity and dissent, they dissent by breaking with the current understanding of structural relationship among constituting components. Their effect throughout the latter is then absorbed and softened by the complex entrenchment of myriads of relationships, forged along the inventive process.

[TABLE 3: CHANGES IN CONVENTIONALITY]

### 4.3. (Un)Conventional Recombinations: Sources and Impact

In the next section, we will analyze the degree of conventionality in inventions as proposed above in a multivariate setting, controlling for some antecedents of the invention under analysis. In particular, we are interested in the role of sources of inventive (un)conventionality as well as the implications of (un)conventionality for subsequent technical developments. As our variables is positively skewed, we use the natural logarithm of it, **Log Conventionality**.[13]

The first dimension we take account of is the extent to which the focal invention builds on existing knowledge. In our setup, we will use the (natural logarithm plus one of) number of citations to prior art as measure of the knowledge base on which the focal invention relies on

(**Log Citations**). We will also differentiate between citations to previous technical literature and scientific literature, that is citations to non-patent literature and include the latter as the share of total citations (**Science**). Furthermore, we include a control for those inventions which do not cite any prior art to account for the possibility that unconventional connections might not find support in any existing knowledge base (**No Prior Art**). We will use the average patent number of the patent documents cited as prior art as measure of the average age of the patent literature which forms the basis of the focal invention (**Age**). Furthermore, we control for the standard deviation of the patent numbers of the patent documents cited as prior art (**Spread Age**). We include also a control for patents citing no patents in the prior art, because for this group we cannot calculate the variable **Age** (**No Patent**) and a control for inventions citing a single patent document as prior art because **Spread Age** cannot be calculated for this group (**Single Citation**). The extent of conventionality embedded in an invention is a positive function of the elements constituting it, that is the components. We hence include the number of patent subclasses the patent has membership in (**Component**).

We further control for some organizational factors affecting the search process. We include the number of inventors comprising the inventive team (**Team**) as well as a measure for single inventor patents (**Single Inventor**). To control for the experience of the inventive team, we include the largest progressive number of patents by the inventors in the team (**Experience**). To account for the organizational determinants of inventive behavior, we have a measure of inventive size of the organizations which appear as assignee on the patent document as the (log plus one) number of patents at USPTO in the year of the focal invention (**Assignee**) as well as a dummy indicating whether the patent was not assigned to any third party and remained to the inventors (**Self**). We finally add **Year** and **Technology** dummies to account for macro trends in the degree of conventionality among patents, such as the introduction of novel patent classes in a given year at USPTO which would artificially alter the measure of recombinant conventionality. The variables are statistically described in Table 4 whereas Table 5 presents bivariate correlations among them.

[TABLE 4: SUMMARY STATISTICS]

[TABLE 5: CORRELATION TABLE]

Table 6 reports the results of OLS. The first column introduces the controls at the level of invention; the second column adds the controls at the level of the inventive team. Finally the last column adds the controls at the level of the assignee. The initial set of controls provide the bulk of the explanatory power, most of which is attributable to year and technology effects: regressing Log Conventionality only on the year and technology on 21 year dummies and 37 technology dummies yields an R-squared of 0.1147. Adding the remaining invention controls improves the explicative power of the model to 0.144. Yet, this improvement is by far the largest when compared to the inclusion of team and assignee level controls.

[TABLE 6: DETERMINANTS OF UNCONVENTIONALITY IN INVENTIONS]

All in all, the results indicate that inventions are more conventional when they embed a limited amount of components and make little use of existing solutions, especially when they are external to the technical domain. Furthermore, conventionality is rooted in familiar and mature solutions which happen to be combined with more recent ones. Inventions being the result of collaborations are less conventional; yet, larger and more experienced teams seem to recombine components in a more conventional fashion. Finally, inventions occurring in larger organizations carry forward unconventional solutions, as opposed to the "garage" inventors, which are conventional in their recombinant efforts.

Conventionality in inventions is negatively associated with the amount of backward citations in patents. A 10% increase in the amount of documents cited as prior art is related to a decrease of 0.2% in the median level of conventionality of the focal invention. Inventions which do not cite any patent as prior art are not more conventional than those which do, whereas inventions which do not cite any prior art show a level of conventionality about 2% higher. Ceteris paribus, the more inventions source from other domains than the technical one – especially from Science – the lower the extent of conventionality in their recombinations: one standard deviation increase in the share of backward citations which are non-patent literature is associated to a 1.25% decrease in the degree of conventionality of the focal invention. As an example, should, everything else equal, all of the 13 documents cited as prior art be patents, our average invention would have a degree of conventionality equals to 40.43. Imagine that, all else equal, now all these documents be citations to scientific references or other sources than patents, than the degree of conventionality of the invention would be on average 38.45. To conclude with the controls at the invention level, the degree of conventionality in recombinations decreases as the number of components used in the focal invention increases: one standard deviation increase in the number of patent subclasses in which the focal invention has memberships in (from 4.6 to almost 8 ) is related to a decrease of 7.68% in the median value of conventionality of the invention, ceteris paribus. Inventions carrying forward conventional recombinations rely to a larger extent on more recent prior art. By the same token, inventions sourcing from more distributed solutions seem to be the result of conventional recombinations. As we measured age by the number of the patent document cited as prior art, both the coefficient of Age and the one of Spread Age are hardly interpretable.

Teams produce inventions with a lower degree of conventionality in the recombinant process as opposed to single inventors. The median value of the extent of conventionality in an invention produced by a single inventor is indeed 4.2% higher. Not only teams are more likely to recombine components unconventionally, this ability grows with the size of the team. By doubling the number of inventors in a team, that is moving from 2 inventors to 4 inventors, conventionality decreases by 0.4% for the average patent. More experienced inventors are more able to combine components in an unconventional fashion.

The final set of controls suggest that larger firms are more likely to be responsible for the generation of inventions which embed unconventional combinations. At the average, doubling the size of the assignee in terms of successful patents applied in a given year decreases the degree of conventionality by 2.1%, all else equal. Furthermore, “garage” inventors, inventors which do not belong to any existing organization and most likely are self-employed, produce more conventional combinations in their inventions. Adding the final set of controls, related to the size and institutional origins of the assignee of the patent, causes some covariates related to the characteristics of the team to change sign: inventor’s experience and team size become positive and significant. We suspect that this has to do with the ability of large firms to attract more experienced inventors and coordinate larger teams.

[TABLE 7: IMPACT OF UNCONVENTIONALITY ON FUTURE INVENTIONS]

Finally, Table 7 presents the results of a set of negative binomial regressions estimating the relationship existing between (un)conventionality and invention's impact. Both Conventionality and Minimum Conventionality are negatively associated with future citations: unconventional combinations are associated with higher impact on future technical developments. However, when they are introduced together in the analysis, Conventionality turns positive and significant, whereas Minimum Conventionality remains negative. The results are confirmed by the inclusion of the interaction term between the two dimensions of Conventionality. All in all, this confirms the evidence from Table 3: unconventionality is associated with higher impact, especially when it is related to the most creative act, as long as it remains embedded in established frameworks.

Impact is positively associated with the number of claims in a patent, the number of backward citations as well as the number of patent classes therein, in line with the view that inventions spanning across wide spectra of the knowledge space will have a higher influence on future inventions. The ratio of citations coming from non-patent literature is negatively associated with impact. This result has to be understood in combination with the coefficient associated to the number of backward citations, indicating that patents drawing mostly from outside the patent literature have a limited impact on future inventions. Finally, inventions from larger teams as well as experienced inventors receive a larger number of future citations, whereas inventions from large applicants receive less citations, ceteris paribus.

## 5. Discussion and Conclusions

In this paper we investigate the origins of unconventional combinations of knowledge components. Unconventional combinations are largely believed to be at the foundation of breakthrough inventions as they establish new connections between distant and overlooked domains of knowledge. In so doing, they remove obstacles and bottlenecks to the combinatorial power of research and development efforts, thus favoring an upsurge of follow on inventions. By considering the inventive process as a process of recombinant search, in our analysis, we first discuss the concept of distance in the search process and how it influences the extent of conventionality in the inventive process. As inventors mostly search locally, they will mostly recombine technological components in a conventional manner, i.e. according to the structure with which relationships have proved to work in the past. By extension, most inventions will be the outcome of conventional combinations.

We thus propose a measure to determine the distance among the elements of the technological space. We borrow the concept and operationalization of relatedness from the literature on product market diversification (Teece et al., 1994) and adapt it to our purpose in the same fashion as in Breschi et al. (2001) and Nesta and Saviotti (2005). We use patent documents at USPTO between 1975 and 2000 and measure conventionality in combinations, and by extension, in inventions at the core of their combinatorial effort and at the most unconventional instance. Our results confirm that most of the recombinant and inventive activities are grounded in conventional efforts, with some rare instances of unconventional connections. Furthermore, we provide suggestive evidence on the relationship between unconventional combinations and future impact. We observe a premium on future impact from unconventionality: inventions embodying conventional combination in their core and carrying forward unconventional combinations in their most unconventional acts are cited more by future patent applications than conventional inventions. Furthermore, we provide indirect evidence that unconventional combinations are at the core of shifts in technological paradigms as they become more conventional over time, suggesting an increase in reuse of the latter. To confirm the goodness of our measure, we identified the main drivers of distance in the search process, which we expected to be responsible for unconventional combinations, and correlated with our measure of unconventionality in inventions and combinations. We find that patents which take a broader view by citing a widespread spectrum of previous results both in science and technology have a higher chance to identify unconventional connections. Moreover, patents having no backward citations of any kind are more conventional. Concluding, unconventionality is more likely to occur with experience, in teams and in large organizations.

The contributions of our work are manifold. From a theoretical standpoint, the results are in line with the body of work on the theory of invention and creativity in general, which posit that agents mostly work in the neighborhood of their competences. Combinations mostly occur with components whose associations have proved to be effective by past use. Inventors eventually experiment with a limited set of components at the time (Fleming, 2001). Much like in Schilling and Greene (2011), this outcome confirms that novel and unconventional combinations are at the origin of high impact solutions as they bridge deep pools of coherent and established knowledge. Unconventional combinations bring together distant concepts and ideas, reshaping the associative framework within which concepts are related and rendering associations that had been overlooked suddenly feasible.

From a methodological standpoint, we are among the firsts to propose a measure to account for the extent of conventionality in the recombinant process, supposedly the ultimate source of novelty. With the exceptions of Fleming (2001) and Dahlin and Behrens (2005), most of the empirical studies on the origins of high-impact inventions have assumed that the ultimate sources of technological impact had to be found in the generation of unconventional combinations and their ability to shape future developments (Ahuja and Lampert, 2001; Fleming and Singh, 2011; Schoenmakers and Duysters, 2011; Kelly et al., 2013). Yet, these studies made no effort to operazionalize this concept. Previous attempts have focused on the very first instance of a combinatorial occurrence (i.e. Fleming et al., 2007; Operti and Carnabuci, 2013). Such approaches operationalize novelty in absolute terms, neglecting the cumulative nature of invention and innovation, whereby novelty is often distributed across early attempts and not constrained to the very first one. This approach is also plagued by a problem of incompleteness, which our approach tries to overcome. To identify absolute novelty, a complete knowledge of all human inventions and the exact time at which they came into existence for the first time is needed. Our measure instead, is population based and hence reflects the state of relationships among the elements of the knowledge space at a given point in time. It only requires to assume that the observed relationships did not vary drastically in the immediate years before they became observable. From Table 4, we learnt that this is not the case as conventionality in combinations tend to change very slowly over time. Attempts to describe unconventionality in idea recombination are in Dahlin and Behrens (2005), Schilling and Greene (2011) and Uzzi et al. (2013). The former study determines conventionality as the overlap in backward citations among patents to determine similarity among patents. This methodology is problematic as the universe of patents is ever expanding and similar inventions might share few backward citations as they occur in two different time periods or because the solution they address is grounded in a multitude of former patents, which might end up not being cited in all the future inventions. Our approach rests on a fairly stable feature of the patent system, the patent classification, which is only marginally subject to variations, and therefore more reliable in the determination of conventionality. Schilling and Greene (2011) use the Dewey decimal system, a bibliographic categorization for the organization of libraries, to determine which combinations of topics is the least likely to occur within the articles cited as references. Their work though is not informative on the actual procedure to determine unconventional connections. A recent study by Uzzi et al. (2013) on the universe of scientific articles in the Web of Science is the closest to ours as they explicitly model novelty in the creative process as the degree of conventionality when domains of knowledge are recombined as reported by the reference lists of the focal articles. As we do, they also take a probabilistic approach as to whether combinations are deterministic or instead the outcome of a random process.

## REFERENCE LIST


Ahuja, G. and Lampert, C.M. (2001). Entrepreneurship in the Large Corporation: A Longitudinal Study of How Established Firms Create Breakthrough Inventions. Strategic Management Journal, 22, pp. 521-543.

Arts S, Veugelers R. (2013). The technological origins and novelty of breakthrough inventions, FEB Research Report - MSI_1302.

Breschi, S., Lissoni, F. and Malerba, F. (2003). Knowledge relatedness in firm technological diversification. Research Policy, 32(1), 69-87

Carnabuci, G. and Operti, E. (2013), Where do firms' recombinant capabilities come from? Intraorganizational networks, knowledge, and firms' ability to innovate through technological recombination. Strat. Mgmt. J., 34: 1591–1613.

Cohen W. and Levinthal D. (1990), Absorptive capacity: A new perspective on learning and innovation", Administrative Science Quarterly, Volume 35, pg. 128-152.

Cyert, R. and March, J. G. (1992). A Behavioral Theory of the Firm (2 ed.). Wiley-Blackwell

Dahlin, K.B. and Behrens, D.M. (2005). When is an invention really radical?: Defining and measuring technological radicalness. Research Policy, 34(5), pp. 717-737.

Dosi, G. (1982), Technological paradigms and technological trajectories. A suggested interpretation of the determinants and directions of technical change, Research Policy, 11(3):147-162.

Dosi, G., (1988). Sources, procedures, and microeconomic effects of innovation. Journal of Economic Literature 26, 1120–1171.

Fleming, L. (2001). Recombinant Uncertainty in Technology Search. Management Science, 47(1), pp. 117-132.

Fleming, L. and Sorenson, O. (2004). Science as a map in technological search. Strategic Management Journal, 25, pp. 909-928.

Fleming, L., Mingo, S. and Chen, D. (2007). Collaborative Brokerage, Generative Creativity and Creative Success. Administrative Science Quarterly, 52(3), pp. 443-475.

Grant, R. M. (1996). Toward a knowledge-base theory of the firm. Strategic Management Journal, 17: 109–122.

Hall B H, Jaffe A, Trajtenberg M. "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools." Working Paper No. 8498, NBER, 2001.

Hargadon A. and Sutton R. (1997). "Technology brokering in a product development firm". Administrative Science Quarterly, 42: 716-749.

Henderson, R. and Clark, K.B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. Administrative Science Quarterly, 35(1), pp. 9-30.

Hill, C. W. L., and Rothaermel, F. T. (2003). The Performance of Incumbent Firms in the Face of Radical Technological Innovation, Academy of Management Review, 28 (2): 257-274.

Katila, R., Ahuja, G., 2002. Something old, something new: a longitudinal study of search behavior and new product introduction. Academy of Management Journal 45, 1183–1194.

Kelley, D. J., Ali, A. and Zahra, S. A. (2013), Where Do Breakthroughs Come From? Characteristics of High-Potential Inventions. Journal of Product Innovation Management, 30: 1212–1226.

Klepper S. and P. Thompson (2010), “Disagreements and Intra-industry Spinoffs,” International Journal of Industrial Organization 28(5), 526-538.

Lai R., D’Amour A., Doolin D., Li G.-C., Sun Y., Torvik V., Yu A. and Fleming L. (2013). Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975-2010). Harvard Business School, Harvard Institute for Quantitative Social Science.

Levinthal D.A., J.G. March (1993). The myopia of learning. Strategic Management Journal, 14, 95-112.

Mullen B., Johnson C. and Salas E. (1991). Productivity Loss in Brainstorming Groups: A Meta-Analytic Integration. Basic and Applied Social Psychology, 12: 3-23.

Neckar, A. (2003). Old is gold? The value of temporal exploration in the creation of new knowledge. Management Science, 49: 211-229.

Nelson, R. and Winter, S. (1982). An Evolutionary Theory of Economic Change. Cambridge, MA: Belknap Press

Nesta, L.;Saviotti, P.P. (2005). Coherence of the knowledge base and the firm's innovative performance : evidence from the U.S. pharmaceutical industry. Journal of Industrial Economics , 53(1) , pp. 123-142.

Paulus, P., B. Nijstad, eds. (2003). Group creativity: An introduction. *Group Creativity: Innovation Through Collaboration*. Oxford University Press, New York.

Polanyi, M (1962) “The Republic of Science: Its Political and Economic Theory”, Minerva, 1, 54-74.

Rosenkopf, L. and Neckar, A. (2001). Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry. Strategic Management Journal, 22 (4), pp. 287–306.

Schilling M, Phelps CC (2007). Interfirm collaboration networks: The impact of large-scale network structure on firm innovation, Management Science, 1113-1126.

Schilling M., Greene E. (2011). Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences. Research Policy, 40 (10), 1321–1331.

Schoenmakers, W. and Duysters, G. (2010). The technological origins of radical inventions. Research Policy, 39, pp. 1051-1059.

Schumpeter J. A. (1939). BUSINESS CYCLES. A Theoretical, Historical and Statistical Analysis of the Capitalist Process. New York Toronto London : McGraw-Hill Book Company, Simonton, D.K., (1999). Creativity as blind variation and selective retention: is the creative process Darwinian? Psychological Inquiry, 10 (4), 309–328.

Singh J, and Fleming L. (2010). Lone Inventors as Sources of Breakthroughs: Myth or Reality?. Management Science, 56: 41-56.

Stuart, T. E. and Podolny, J. M. (1996), Local search and the evolution of technological capabilities. Strat. Mgmt. J., 17: 21–38.

Teece DJ, R Rumelt, G Dosi, S Winter (1994). Understanding corporate coherence: theory and evidence. Journal of Economic Behavior and Organization 23: 1-30

Uzzi B., Mukherjee S., Stringer M., Jones B. (2013). Atypical Combinations and Scientific Impact. Science, 342, 468-472.

Van de Poel, I. (2003). The transformation of technological regimes. Research Policy, 32(1), 49-68

Weitzman, M. L. (1998). Recombinant growth. Quarterly Journal of Economics 113(2): 331-360.

## List of Tables and Figures

### Table 1: Distribution of Conventionality across Years

| Year | Conventionality | Std. Dev. | Occurrences | Year | Conventionality | Std. Dev. | Occurrences |
|---|---|---|---|---|---|---|---|
| 1980 | 52,323 | 43,133 | 57185 | 1991 | 45,385 | 41,392 | 90331 |
| 1981 | 51,4 | 42,716 | 55584 | 1992 | 44,17 | 40,281 | 93781 |
| 1982 | 51,431 | 42,669 | 56723 | 1993 | 44,164 | 41,12 | 97664 |
| 1983 | 50,915 | 43,079 | 54310 | 1994 | 44,067 | 41,128 | 111428 |
| 1984 | 51,028 | 42,623 | 59401 | 1995 | 44,039 | 41,367 | 130686 |
| 1985 | 50,134 | 42,718 | 63264 | 1996 | 43,079 | 43,015 | 129961 |
| 1986 | 49,412 | 41,448 | 66885 | 1997 | 43,315 | 43,585 | 152371 |
| 1987 | 48,885 | 41,995 | 72710 | 1998 | 42,328 | 44,759 | 151632 |
| 1988 | 48,056 | 41,972 | 80404 | 1999 | 42,087 | 44,175 | 161870 |
| 1989 | 47,302 | 41,45 | 85728 | 2000 | 43,551 | 47,141 | 176747 |
| 1990 | 46,47 | 41,802 | 89066 | Total | 47,535 | 47,6 | 2037731 |

## Table 2: Distribution of Conventionality across Technology fields

| Technology | Conventionality | Std. Dev. | Occurrences |
|---|---|---|---|
| Agriculture, Food, Textiles | 53,115 | 45,144 | 11560 |
| Agriculture, Husbandry, Food | 60,601 | 59,772 | 34506 |
| Amusement Devices | 73,599 | 58,319 | 16894 |
| Apparel and Textile | 79,359 | 7,064 | 26113 |
| Biotechnology | 82,237 | 78,216 | 4063 |
| Coating | 36,175 | 33,920 | 31443 |
| Communications | 36,042 | 32,303 | 135267 |
| Computer Hardware and Software | 35,834 | 32,876 | 97670 |
| Computer Peripherals | 32,789 | 29,690 | 39590 |
| Drugs | 32,523 | 28,452 | 122574 |
| Earth Working and Wells | 61,312 | 53,306 | 24246 |
| Electrical Devices | 47,313 | 43,925 | 60206 |
| Electrical Lighting | 43,223 | 33,652 | 34051 |
| Furniture, House Fixtures | 70,169 | 55,830 | 37354 |
| Gas | 53,399 | 40,156 | 8644 |
| Heating | 54,829 | 50,558 | 21891 |
| Information Storage | 31,988 | 30,950 | 62151 |
| Materials Processing. and Handlin | 53,687 | 45,190 | 91551 |
| Measuring and Testing | 45,199 | 39,677 | 50698 |
| Metal Working | 53,381 | 46,162 | 53466 |
| Miscellaneous – Drug and Med | 60,126 | 53,105 | 11373 |
| Miscellaneous-Electrical | 43,585 | 3,673 | 65037 |
| Miscellaneous-Mechanical | 63,634 | 54,843 | 82170 |
| Miscellaneous-Others | 46,518 | 45,921 | 189501 |

| | | | |
|---|---|---|---|
| Miscellaneous-chemical | 41,772 | 36,188 | 190822 |
| Motors, Engines and Parts | 57,402 | 49,889 | 65460 |
| Nuclear and X-rays | 40,889 | 35,031 | 29305 |
| Optics | 45,724 | 41,370 | 19241 |
| Organic Compounds | 51,072 | 46,105 | 46807 |
| Pipes and Joints | 44,285 | 35,655 | 15850 |
| Power Systems | 44,288 | 38,253 | 70136 |
| Receptacles | 48,085 | 36,214 | 34402 |
| Resins | 29,178 | 24,073 | 68225 |
| Semiconductor Devices | 31,521 | 0,245 | 71551 |
| Surgery and Medical Instruments | 42,099 | 36,098 | 62261 |
| Transportation | 70,258 | 61,359 | 51652 |
| Total | 45,639 | 43,012 | 2037731 |

**Table 3: Changes in Conventionality: Fixed-Effect panel data regression – Whole sample and sample split by initial degree of initial conventionality**

| **Conventionality** | **All** | **10th centile** | **10 to 25th** | **25 to 50th** | **50 to 75th** | **75 to 90th** | **Top 10th** |
|---|---|---|---|---|---|---|---|
| Time trend | -0,236 | 0,053 | -0,019 | -0,006 | -0,163 | -0,266 | 0,106 |
| Constant | 41,339 | 5,108 | 10,364 | 19,259 | 35,730 | 65,055 | 168,853 |
| Time dummies | YES | YES | YES | YES | YES | YES | YES |
| Observations | 20891560 | 1521625 | 3245670 | 5642931 | 5513696 | 3182778 | 1759495 |

**Table 4: Summary Statistics split by degree of Median Conventionality ($10^{th}$ of least Conventional inventions)**

| | Full Sampe | | | 90% Most Conventional | | | 10% Least Conventional | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **Obs** | **Mean** | **Std. Dev.** | **Obs** | **Mean** | **Std. Dev.** | **Obs** | **Mean** | **Std. Dev.** | **T-test** |
| Log Conv | 2037731 | 3,706 | 0,65 | 1833958 | 3,821 | 0,575 | 203773 | 2,677 | 0,196 | +++ |
| Citations (log) | 2037731 | 2,199 | 0,83 | 1833958 | 2,185 | 0,816 | 203773 | 2,327 | 0,910 | --- |
| Science | 2037731 | 0,139 | 0,26 | 1833958 | 0,133 | 0,252 | 203773 | 0,189 | 0,288 | --- |
| Component | 2037731 | 4,659 | 3,27 | 1833958 | 4,608 | 3,213 | 203773 | 5,122 | 3,692 | --- |
| Age | 2037731 | 63,569 | 4,78 | 1833958 | 59,572 | 46,209 | 203773 | 99,542 | 60,432 | --- |
| Spread Age | 2037731 | 102,384 | 77454,44 | 1833958 | 99,486 | 81,510 | 203773 | 128,466 | 14026,150 | |
| No Patent | 2037731 | 0,028 | 0,16 | 1833958 | 0,027 | 0,161 | 203773 | 0,036 | 0,185 | --- |
| No Prior Art | 2037731 | 0,012 | 0,11 | 1833958 | 0,012 | 0,110 | 203773 | 0,013 | 0,113 | --- |
| Team | 2037731 | 2,216 | 1,57 | 1833958 | 2,187 | 1,556 | 203773 | 2,484 | 1,696 | --- |
| Experience | 2037731 | 11,681 | 27,56 | 1833958 | 11,330 | 26,749 | 203773 | 14,838 | 33,873 | --- |
| Single Inventor | 2037731 | 0,424 | 0,49 | 1833958 | 0,433 | 0,496 | 203773 | 0,337 | 0,473 | +++ |
| Assignee (log) | 2037731 | 3,540 | 2,76 | 1833958 | 3,729 | 2,740 | 203773 | 1,834 | 0,000 | --- |
| Self | 2037731 | 0,153 | 0,36 | 1833958 | 0,162 | 0,369 | 203773 | 0,074 | 0,262 | +++ |
| No Age | 2037731 | 0,000 | 0,01 | 1833958 | 0,000 | 0,014 | 203773 | 0,000 | 0,015 | |
| Single Patent | 2037731 | 0,033 | 0,18 | 1833958 | 0,034 | 0,180 | 203773 | 0,027 | 0,163 | --- |

**Table 5: Correlation Table**

| | Log Conv | Cits | Science | Compo | Age | Spread Age | No Patent | No Prior Art | Team | Exp | Single Inventor | Assignee |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Log Conv | | | | | | | | | | | | |
| Citations | -0.084* | | | | | | | | | | | |
| Science | -0.103* | 0.352* | | | | | | | | | | |
| Component | -0.198* | 0.112* | 0.121* | | | | | | | | | |
| Age | -0.003* | 0.006* | 0.001 | 0.001 | | | | | | | | |
| Spread Age | -0.001 | 0.004* | 0.001 | 0.000 | 0.919* | | | | | | | |
| No Patent | -0.011* | -0.213* | 0.272* | 0.03* | -0.002* | -0.001 | | | | | | |
| No Prior Art | 0.012* | -0.297* | -0.060* | -0.001 | -0.001 | -0.001 | 0.663* | | | | | |
| Team | -0.099* | 0.074* | 0.146* | 0.104* | 0.001 | 0.001 | 0.053* | 0.015* | | | | |
| Experience | -0.059* | 0.042* | 0.029* | 0.064* | 0.001 | 0.001 | 0.015* | 0.009* | 0.193* | | | |
| Single Inventor | 0.106* | -0.071* | -0.144* | -0.088* | -0.002* | -0.001 | -0.043* | -0.008* | -0.663* | -0.142* | | |
| Assignee | -0.201* | 0.006* | 0.110* | 0.068* | 0.002* | -0.001 | -0.010* | -0.031* | 0.282* | 0.194* | -0.291* | |
| Self | 0.142* | -0.051* | -0.128* | -0.065* | -0.001 | -0.001 | 0.011* | 0.052* | -0.235* | -0.093* | 0.301* | -0.592* |

**Table 6: Determinants of Conventionality**

| | OLS | | | QUANTILE ($10^{th}$ Centile) | | | LOGIT (10th Centile) | | |
|---|---|---|---|---|---|---|---|---|---|
| Log Citations | -0.018*** | -0.017*** | -0.021*** | -0.011*** | -0.010*** | -0.014*** | 0.030*** | 0.026*** | 0.040*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.003) | (0.003) | (0.003) |
| Science | -0.025*** | -0.020*** | -0.005** | -0.051*** | -0.046*** | -0.035*** | 0.168*** | 0.156*** | 0.119*** |
| | (0.002) | (0.002) | (0.002) | (0.003) | (0.003) | (0.003) | (0.012) | (0.012) | (0.012) |
| Component | -0.024*** | -0.023*** | -0.023*** | -0.001*** | -0.001*** | -0.001*** | 0.017*** | 0.016*** | 0.015*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.001) | (0.001) |
| Age | -0.000** | -0.000** | -0.000** | -0.000 | -0.000 | -0.000 | 0.000* | 0.000* | 0.000* |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Spread Age | 0.000** | 0.000** | 0.000** | 0.000 | 0.000 | 0.000 | -0.000 | -0.000 | -0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| No Patent | 0.006 | 0.005 | -0.007 | -0.005 | -0.005 | -0.014*** | -0.023 | -0.023 | 0.005 |
| | (0.004) | (0.004) | (0.004) | (0.005) | (0.005) | (0.005) | (0.019) | (0.019) | (0.019) |
| No Prior Art | 0.020*** | 0.027*** | 0.010* | -0.006 | 0.000 | -0.017** | 0.052* | 0.036 | 0.145*** |
| | (0.006) | (0.006) | (0.006) | (0.007) | (0.007) | (0.007) | (0.029) | (0.029) | (0.029) |
| Team | | -0.002*** | 0.004*** | | -0.001* | 0.003*** | | -0.007*** | -0.022*** |
| | | (0.000) | (0.000) | | (0.000) | (0.000) | | (0.002) | (0.002) |
| Experience | | -0.000*** | 0.000*** | | -0.000*** | 0.000 | | 0.000*** | -0.000*** |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | (0.000) | (0.000) | | (0.000) | (0.000) | | (0.000) | (0.000) |
| Single Inventor | | 0.042*** | 0.021*** | | 0.034*** | 0.019*** | | -0.150*** | -0.084*** |
| | | (0.001) | (0.001) | | (0.002) | (0.002) | | (0.007) | (0.007) |
| Assignee | | | -0.021*** | | | -0.017*** | | | 0.061*** |
| | | | (0.000) | | | (0.000) | | | (0.001) |
| Self | | | 0.025*** | | | 0.026*** | | | -0.198*** |
| | | | (0.002) | | | (0.002) | | | (0.011) |
| Constant | 4.148*** | 4.135*** | 4.211*** | 3.342*** | 3.330*** | 3.395*** | -4.909*** | -4.830*** | -5.047*** |
| | (0.006) | (0.006) | (0.006) | (0.009) | (0.009) | (0.009) | (0.072) | (0.072) | (0.072) |
| N | 2037731 | 2037731 | 2037731 | 2037731 | 2037731 | 2037731 | 2037731 | 2037731 | 2037731 |
| R-sq | 0.144 | 0.146 | 0.153 | | | | | | |

***, **, and *** indicate respectively 10%, 5% and 1% statistical significance.** The first three columns reports the results of Ordinary Least Square on the median value of conventionality in patents. The second set of columns instead report the results of Quantile regressions at the bottom 10th centile. The last set of columns report the results of a logit regressions on the likelihood of a patent of belonging to the most unconventional 10%. Regressions include 21 year dummies and 37 technology dummies; all dummies are jointly statistically significant. Regressions include also controls (dummies) for missing information concerning the age of the backward citations and whether the backward citations is made of one single patent. Standard Errors are robust to outliers in the case of the OLS results in the first three columns.

**Table 7: Implications of Conventionality for Future Impact**

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Conv | -0.0771*** | | 0.0420*** | 0.0685*** |
| | (0.0017) | | (0.0035) | (0.0069) |
| Min Conv | | -0.1030*** | -0.1394*** | -0.1002*** |
| | | (0.0018) | (0.0036) | (0.0091) |
| Conv*Min Conv | | | | -0.0088*** |
| | | | | (0.0019) |
| Log Claim | 0.0133*** | 0.0133*** | 0.0133*** | 0.0133*** |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| Log Citations | 0.2249*** | 0.2231*** | 0.2228*** | 0.2228*** |
| | (0.0015) | (0.0015) | (0.0015) | (0.0015) |
| Science | -0.0231*** | -0.0233*** | -0.0236*** | -0.0233*** |
| | (0.0054) | (0.0054) | (0.0054) | (0.0054) |
| Component | 0.0332*** | 0.0248*** | 0.0224*** | 0.0228*** |
| | (0.0004) | (0.0004) | (0.0004) | (0.0004) |
| Age | -0.0000* | -0.0000* | -0.0000* | -0.0000* |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| Spread Age | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

| | | | | |
|---|---|---|---|---|
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| No Patent | -0.1181*** | -0.1230*** | -0.1245*** | -0.1245*** |
| | (0.0159) | (0.0156) | (0.0155) | (0.0155) |
| No Prior Art | 0.3956*** | 0.3993*** | 0.4007*** | 0.4009*** |
| | (0.0196) | (0.0194) | (0.0193) | (0.0193) |
| Team | 0.0309*** | 0.0311*** | 0.0311*** | 0.0311*** |
| | (0.0009) | (0.0009) | (0.0009) | (0.0009) |
| Experience | 0.0001* | 0.0001* | 0.0001* | 0.0001* |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| Single Inventor | -0.0433*** | -0.0425*** | -0.0425*** | -0.0425*** |
| | (0.0027) | (0.0027) | (0.0027) | (0.0027) |
| Assignee | -0.0075*** | -0.0078*** | -0.0077*** | -0.0077*** |
| | (0.0005) | (0.0005) | (0.0005) | (0.0005) |
| Self | -0.0519*** | -0.0520*** | -0.0526*** | -0.0526*** |
| | (0.0033) | (0.0033) | (0.0033) | (0.0033) |
| Constant | 1.2918*** | 1.4004*** | 1.3747*** | 1.2540*** |
| | (0.0182) | (0.0182) | (0.0183) | (0.0321) |
| N | 2037026 | 2037026 | 2037026 | 2037026 |

***, **, and *** indicate respectively 10%, 5% and 1% statistical significance.** Regressions include 21 year dummies and 37 technology dummies; all dummies are jointly statistically significant. Regressions include also controls (dummies) for missing information concerning the age of the backward citations and whether the backward citations is made of one single patent. The log-transformed over-dispersion parameter, unreported, is always significantly different from zero.

**Figure 1: Cumulative distribution of (Log of) Conventionality in combinations.**

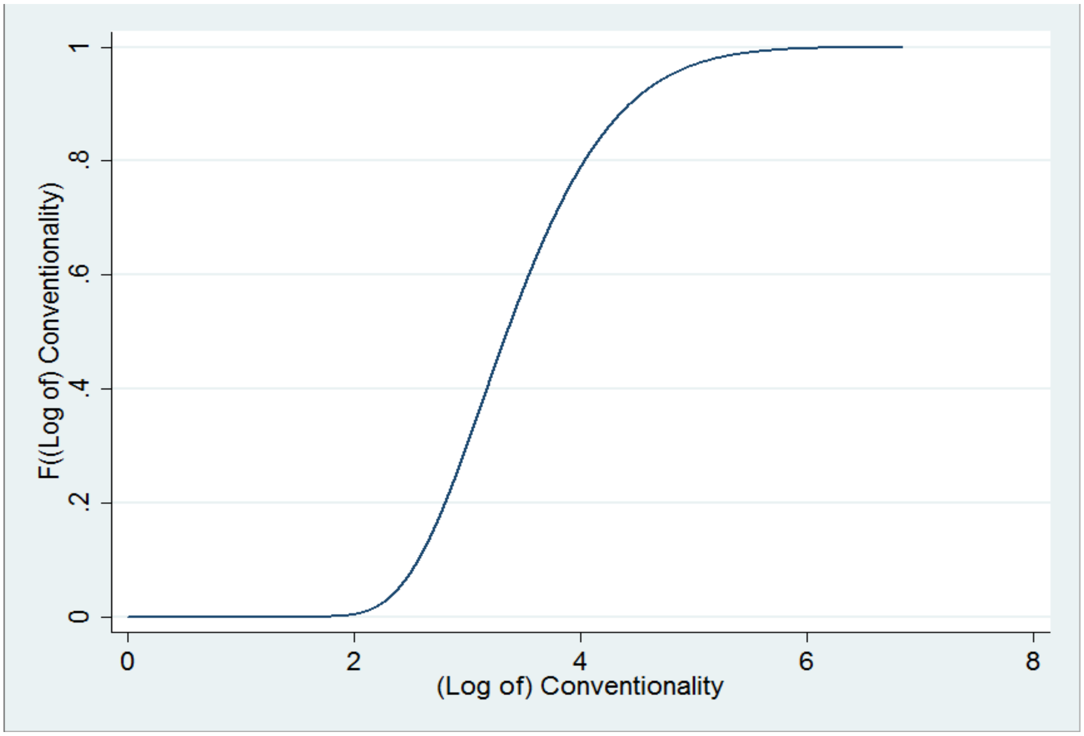
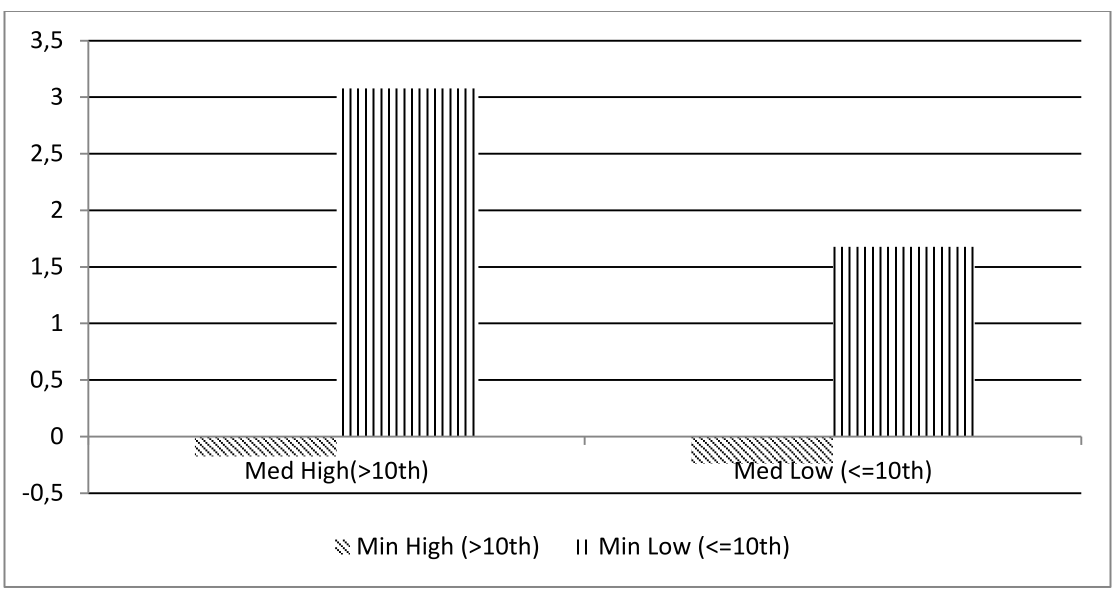



**Figure 2: Forward Citations by Extent of Conventionality**




**Med High (Med Low)** refers to patents whose median conventionality is above (below) the 10th centile of the distribution of median conventionality. Similarly, **Min High (Min Low)** refers to patents whose minimum conventionality is above (below) the $10^{th}$ centile of the distribution of minimum conventionality.

## Appendix: Analytical derivation of Unconventionality in Recombinations

Teece et al. (1994) have developed measures of relatedness and coherence for the diversification activities of firms. In the present study these measures are adapted to describe the diversification patterns in the knowledge space (Breschi et al., 2003; Nesta and Saviotti, 2005; Piscitello, 2005).

Following Teece et al. (1994), let $C_{ik} = 1$ if invention $k$ has membership in patent class $i$, and 0 otherwise. The number of inventions with membership in class $i$ is $n_i = \sum_k C_{ik}$, and the number of inventions having simultaneously membership in both classes $i$ and $j$ is $J_{ij} = \sum_k C_{ik} C_{jk}$.

Raw counts of the number of inventions having membership in each couple of patent classes, however, cannot be taken directly as a measure of relatedness. Classes must be present at a rate greater than what one would expect if combinations were made at random. Although $J_{ij}$ increases with the relatedness of $i$ and $j$, it also increases with $n_i$ and $n_j$, the number of inventions having membership in each class of the couple. Therefore, $J_{ij}$ must be adjusted for the number of inventions that would appear in the couple $ij$ under the null hypothesis that inventors combine patent classes at random.

To operationalize the null hypothesis, the distribution of $J_{ij}$ must be derived by assuming that inventions are assigned to classes at random. For now, call this random variable $x_{ij}$. Teece et al. (1994) identify the distribution of the random variable, but they do not derive it in their paper. For the sake of exposition, we will to derive the distribution in order to clarify the construction of the measure. This brief exposition is similar to Bryce and Winter (2006).

Draw a sample of size $n_i$ from the population of $K$ multi-class inventions. Now draw another sample of size $n_j$ and observe $x_{ij}$, or the number of inventions that were also in the $n_i$ sample. The number of ways of selecting $x$ inventions to fill $x$ positions in sample $n_j$ is equivalent to the number of ways of selecting $x$ from a total of $n_i$ inventors, or $\binom{n_i}{x}$. The number of ways of selecting inventors not receiving assignment to class $i$ for the remaining $(n_j - x)$ positions in the $n_j$ sample is equivalent to the number of ways of selecting $(n_j - x)$ from a possible $(K - n_i)$ inventors, or $\binom{K-n_i}{n_j-x}$. Then the number of possible permutations of the $n_j$ sample is the number of ways of combining a set of $x$ inventions assigned to class $i$ $(n_i)$ multiplied by $(n_j - x)$ inventions not assigned to class $i$, or $\binom{n_i}{x}\binom{K-n_i}{n_j-x}$.[14] The number of different samples of size $n_j$ that can be drawn from $K$ is $\left(\frac{K}{n_j}\right)$. The number of possible permutations of the $n_j$ sample divided by the number of ways of choosing a sample of size $n_j$ is the probability that $x$ inventions from population $K$ are assigned to both class $i$ and class $j$. Thus, the number $x_{ij}$ of inventions having membership in both class $i$ and class $j$ is a hypergeometric random variable

$$P\left[X_{ij} = x\right] = \frac{\binom{n_i}{x}\binom{K-n_i}{n_j-x}}{\binom{K}{n_j}} \quad . \tag{1}$$

whose mean of $X_{ij}$ is [15]

$$\mu_{ij} = E(X_{ij}) = \frac{n_i n_j}{K} \tag{2}$$

and the variance of $X_{ij}$ is

$$\sigma_{ij}^2 = \mu_{ij}\left(1-\frac{n_i}{K}\right)\left(\frac{K-n_j}{K-1}\right). \tag{3}$$

The difference between $J_{ij}$ and the expected value of the random variable $n_{ij}$, or $\tau_{ij}$ , is standardized as

$$\tau_{ij} = \frac{J_{ij}-\mu_{ij}}{\sigma_{ij}}. \tag{4}$$

And forms the basis for the measure of conventionality in combinations. When this difference is positive and large, it indicates that the combination of pairs of patent classes in multi-class inventions is systematic, typical or conventional. When it is negative, it indicates that unexpectedly few inventions have successfully combined the focal couple, suggesting that the combination thereof is not systematic, unconventional or unconventional.

From (4), we can derive the degree of conventionality of the patent $z$, $a_z$, as the simple average of the measure $\tau_{ij}$ for all combinations of technologies ($i$,$j$) whose the patent has membership.

$$InventionConventionality = a_z = \frac{1}{n}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\tau_{ij}\ , \tag{5}$$

Where $n$ is the number of the patent's subclass combinations and $m$ is the combination index. For instance, if a patent has four subclasses, then $m$ is equal to six, since this is the number of subclass combinations (4(4-1)/2). Hence, $m$=1, …, 6.

[1] We thank Ludovic Dibiaggio, Gino Cattani, Jian Wang and participants to the KTO Workshop (Sophia Antipolis June 2013) for useful comments on a previous version of this article. The current version has benefited from informal discussions with department members at MSI-KU Leuven and LIME-IMT Lucca. Timon Gaertner provided useful research assistance.

[2] Jung and Lee (2013) report different definitions of the components involved in the recombinant process employed in the literature. Components are considered as “conceptual or physical materials”, such as routines or technologies (Nelson and Winter, 1982); “old knowledge,” such as existing cultivated plant varieties (Weitzman, 1998); pre-existing “elements,” such as materials in periodic tables, and “conditions,” such as temperature and pressure (Romer, 1994); and “constituents of invention,” such as Schumpeterian “factors” (Schumpeter, 1939; Fleming, 2001).

[3] The psychological literature has also stressed that newer, and thus more creative, combinations are those which are apparently not related among each other. Simonton (1999) pointed out that many of the most famous scientific breakthroughs occurred through a free associative process (what Freudians might call “primary process thinking”). Agents generate many unusual combinations between different bodies of knowledge that set to a screening process of selective retention, keeping only the best variations (much like Darwinian evolution).

[4] By extension, we can think of the degree of relatedness between two components of the knowledge space as the strength of the link between them. Like in the parallel of knowledge or technological landscapes (Fleming, 2001), coherent areas of the knowledge networks are made of highly interrelated components, where the use of one component is usually associated to the use of other ones. Alternatively, there will be combinations of components which link otherwise disconnected areas; these links will be weaker, or less related, than the tighter ones characterizing the coherent sections of the knowledge space. Consequently, the knowledge space can be thought as a network, made of areas of highly interrelated components, eventually connected by unconventional or unconventional combinations (Shilling and Greene, 2011).

[5] Details on the derivation of the measure and formulae to calculate $\mu_{ij}$ and $\sigma_{ij}$ are reported in the appendix.

[6] The index of relatedness $\tau_{ij}$ can also be interpreted as the centripetal strength that ties together the nodes (patent subclasses) of the cognitive space in which inventions occur. High values indicate that two elements are very close in space or interdependent as in Fleming (2001). Intuitively, components which are largely used – large $n_i$ – are indeed hardly interdependent with other components.

[7] A patent is a legal instrument that protects a new and useful product, process, machine, or new combinations of materials. Patents are especially useful for analyzing inventions because they are granted only to products and processes that a knowledgeable, objective third party (e.g. United States Patent and Trademark Office USPTO) decides that the work exceeds a minimum threshold of creativity and innovation.

[8] As $\tau_{ij}$ can take negative values, we summed it by its minimum value, and took the natural logarithm plus 1.

[9] Subject matter directed to self-replicating nucleic acid molecules which may be employed to introduce a nucleic acid sequence or gene into a cell; such nucleic acid molecules are designated as vectors and may be in the form of a plasmid, hybrid plasmid, cosmid, viral vector, bacteriophage vector, etc.

[10] Subject matter directed to methods of searching for (i.e., querying) data stored as a database in a computer or digital data processing system, including sequential searching, primary and secondary index searching, and bit-map searching of inverted lists or topological maps.

[11] We have first calculated the expected number of forward citations as function of the number of claims, year and technology dummies and then subtracted them to the number of actual citations. Positive values indicate that inventions receive more citations than otherwise predicted.

[12] When inventions embody unconventional combinations in their core, but seem fairly typical in their most novel solutions (lower right quadrant) receive the least amount of forward citations.

[13] As our measure of conventionality takes negative values, we added the absolute of the lowest value taken by *Conventionality*. We then took the natural logarithm of the newly transformed covariate plus one.

[14] Since sample $n_j$ was fixed as the number of inventions in class $j$, inventions assigned to class $i$ in this quantity are *de facto* also assigned to class $j$.

[15] Intuition for the mean of (1) is as follows. Assume that $n_j$ inventions in $K$ have been assigned to class $j$. Now randomly assign inventions in $K$ to class $i$. The probability that any one invention receives a class $i$ assignment is $\frac{n_i}{K}$. Since there are $n_j$ inventions in $K$, each with probability $\frac{n_i}{K}$ of being assigned to class $i$, the expected number of inventions assigned to both class $i$ and class $j$ is $n_j\left(\frac{n_i}{K}\right)$.

**FACULTY OF ECONOMICS AND BUSINESS**
**DEPARTMENT OF MANAGERIAL ECONOMICS, STRATEGY AND INNOVATION**
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 67 00
fax + 32 16 32 67 32
info@econ.kuleuven.be
www.econ.kuleuven.be/MSI